

08/02/12

אוניברסיטת תל אביב – ביה"ס למדעי המחשב

גנומיקה חישובית 0382.3102.01 -

תשע"ב

מרצים: פרופ' רון שמיר, פרופ' רודד שרן

מתרגל: ירון אורנשטיין

משך הבחינה: שלוש שעות ללא אפשרות הארכה

חומר עזר: שני דפים כתובים בכתב יד (משני הצדדים)

יש לענות על שלוש מתוך ארבע השאלות. לכל השאלות ניקוד שווה.

יש לענות בצורה קצרה, ברורה ומדויקת ולנמק כל טיעון. תשובה לא מנומקת לא תזכה בנקודות.

בשאלות בהן נדרש לתאר אלגוריתם יש לציין במפורש את סיבוכיות הזמן.

בשאלות בהן השתמשת באלגוריתם שתואר בשיעור או בתרגיל אין צורך לחזור עליו לפרטיו אלא רק לפרט במדויק את השינויים לעומת מה שנלמד.

שאלה 1

עבור מספר שלם $k > 1$, נאמר שמחרוזת s באורך n היא k -מחזורית אם יש מחרוזת u באורך $n/2 \geq$, כך ש- $s = u^k$, u^k היא שרשור של k העתקים של u ו- u היא רישא של u . אורך המחזור הוא אורך המחזור u .

תארי/ אלגוריתם יעיל ככל האפשר המוצא את אורך המחזור הקצר ביותר של s או מחליט ש- s אינה מחזורית.

שאלה 2

כידוע ה-rna הראשוני של גן אאוקריוטי מורכב מאקסונים שביניהם אינטרונים. בתהליך שיחבור רגיל משורשרים האקסונים לפי סדרם ויוצרים יחד את ה-mRNA (הבוגר). בתופעה הנקראת שחבור אלטרנטיבי מדלג תהליך השחבור מעל אקסונים מסויימים וכתוצאה מכך הם אינם מוכלים ב-mRNA הבוגר. תארי/ במלואו אלגוריתם יעיל ככל האפשר המקבל את רצפי כל האקסונים לפי סדרם ורצף מרנא בוגר ומזהה מהם האקסונים המוכלים בו. בשל שגיאות ריצוף ייתכנו אי התאמות בין רצף ה-mRNA הבוגר המתאים לאקסון ורצף האקסון עצמו.

השתמש/י בסימונים הבאים: רצף המרנא הבוגר: $M(1, \dots, m)$. הרצף של האקסונים $E(1, \dots, n)$. נקודות ההתחלה של האקסונים על פני הרצף: $s_1 = 1 < s_2 < \dots < s_k < n$. היינו: אקסון i הוא $E(s_i), \dots, E(s_{i+1}) - 1$. ציון $\sigma(x, y)$: ציון indel: δ . בחישוב הסיבוכיות התייחס/י לכך ש- $k \ll n$.

www-11

שאלה 3

עליך ללמוד מודל מרקובי נסתר לזיהוי איי CpG ברצף ארוך x באורך L .

א. (13 נק') רשום/מי את המודל ההסתברותי כולל דיאגרמת המצבים והמעברים ביניהם וציון הפרמטרים שיש ללמוד.

ב. (10 נק') רשום/מי את פונקציית הנראות של המודל ואת הפונקציה Q המתאימה (תוחלת לוג הנראות השלמה) בפירוט רב ככל האפשר.

ג. (10 נק') תאר/י בפרוטרוט את אלגוריתם באום-וולש ללימוד פרמטרי המודל (אין צורך להוכיח את הנוסחאות).

שאלה 4

נתונה מטריצת דמיון בין n גנים המבוססת על נתוני ביטוי שלהם בתנאים שונים. בהפעלת אלגוריתם קיבוץ (clustering) הררכי על נתונים אלו התקבל עץ בינארי שעליך להציג למשתמש. החופש בהצגה הוא בסידור העלים (היינו הגנים); הסדר חייב להציג ברציפות את העלים של כל תת עץ מושרש. ישנם 2^{n-1} סידורים לינאריים אפשריים). תן/י אלגוריתם יעיל ככל האפשר למציאת סידור בו סכום $n-1$ ערכי הדמיון בין עלים סמוכים הוא מירבי.

בהצלחה!